

## PATENT ABSTRACTS OF JAPAN

(11)Publication number : 2000-003200

(43)Date of publication of application : 07.01.2000

(51)Int.Cl.

G10L 21/04

G10K 15/04

G10L 13/00

(21)Application number : 10-185691

(71)Applicant : YAMAHA CORP

(22)Date of filing : 16.06.1998

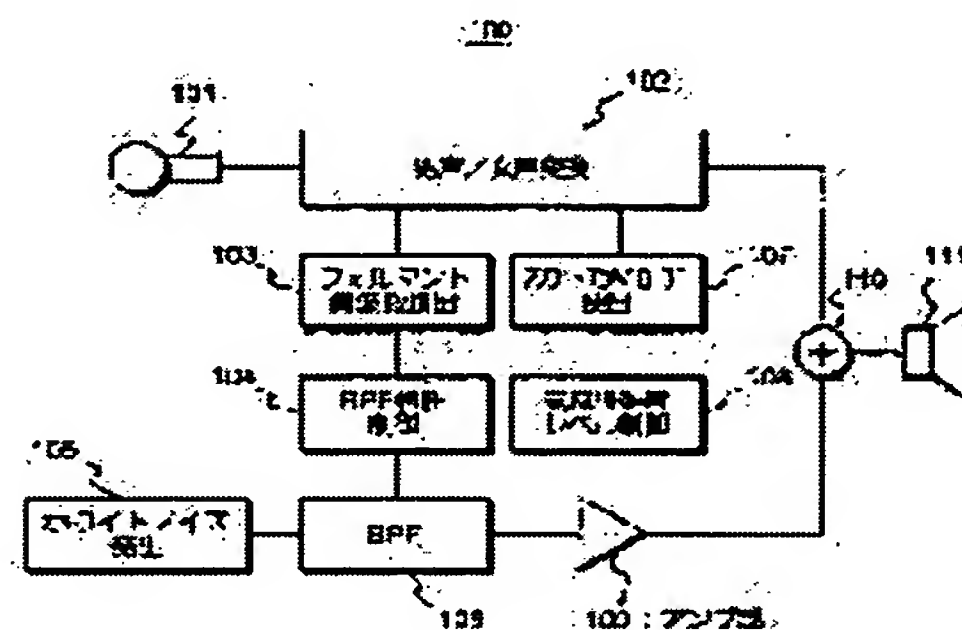
(72)Inventor : KONDO TAKAYASU  
XAVIER SERA

## (54) VOICE SIGNAL PROCESSOR AND VOICE SIGNAL PROCESSING METHOD

## (57)Abstract:

PROBLEM TO BE SOLVED: To easily obtain a female converted voice natural on hearing sense by adding an aspirated noise component signal to an original conversion voice signal and outputting it as a conversion voice signal.

SOLUTION: A band-pass filter part 106 passes only a white noise signal having a frequency in a prescribed frequency band corresponding to the third formant of the original conversion voice signal detected by a formant frequency detection part 103 among the white noise signals outputted by a white noise generation part 105 under control of a band-pass filter characteristic control part 104 through as an original aspirated noise signal to output it to an amplifier part 109. An aspirated noise level control part 108 outputs the level control signal of the original aspirated noise signal to the amplifier part 109 based on the size of the amplification of the original conversion voice signal detected by an amp/envelope detection part 107. The amplifier part 109 changes the signal level of the original aspirated noise signal based on this signal to output it to a mixer part 110 as an aspirated noise signal.



## LEGAL STATUS

[Date of request for examination] 09.03.2000

[Date of sending the examiner's decision of rejection]

[Kind of final disposal of application other than the examiner's decision of rejection or application converted registration]

[Date of final disposal for application]

[Patent number] 3502268

[Date of registration] 12.12.2003

[Number of appeal against examiner's decision]

(19) 日本国特許庁 (J P)

(12) 公開特許公報 (A)

(11)特許出願公開番号

特開2000-3200

(P2000-3200A)

(43)公開日 平成12年1月7日(2000.1.7)

(51) Int.Cl. <sup>7</sup>	識別記号	F I	テマコード*(参考)
G 1 0 L 21/04		G 1 0 L 3/02	A 5 D 0 4 5
G 1 0 K 15/04	3 0 2	G 1 0 K 15/04	3 0 2 D 5 D 1 0 8
G 1 0 L 13/00		G 1 0 L 9/02	L

審査請求 未請求 請求項の数 8 F D (全 15 頁)

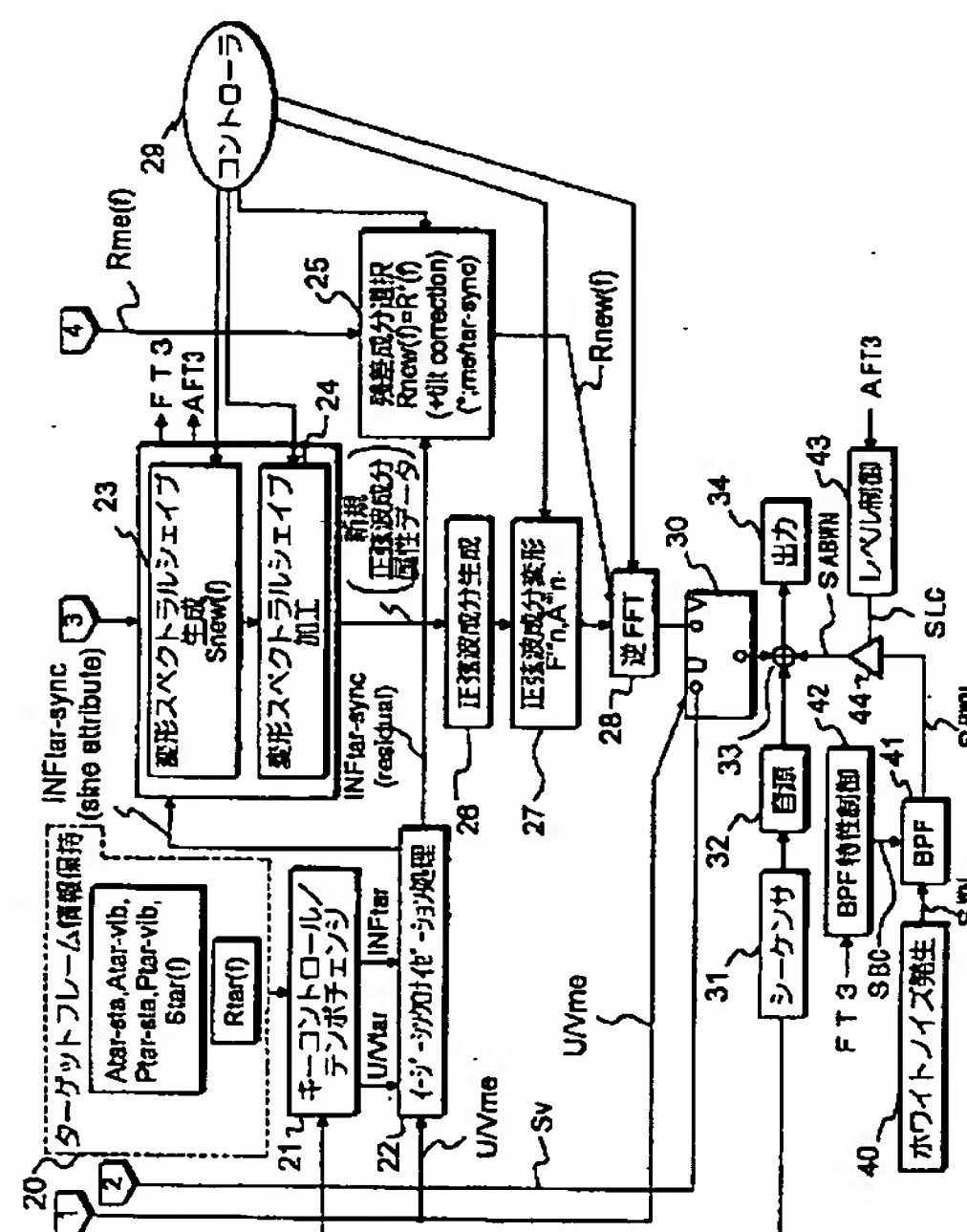
(21)出願番号	特願平10-185691	(71)出願人	000004075 ヤマハ株式会社 静岡県浜松市中沢町10番1号
(22)出願日	平成10年6月16日(1998.6.16)	(72)発明者	近藤 高康 静岡県浜松市中沢町10番1号 ヤマハ株式 会社内
		(72)発明者	ザビエル セラ スペイン パルセロナ カルデデュー 08440 2-2 ビスカイア19
		(74)代理人	100098084 弁理士 川▲崎▼ 研二 (外1名)
		Fターム(参考)	5D045 BA01 5D108 BF02

(54) 【発明の名称】 音声信号処理装置及び音声信号処理方法

(57) 【要約】

【課題】 男性歌唱者の音声を女声の音声に変換するに際し、あるいは、女声の音声を合成するに際し、より自然な女性の歌声に近い出力音声を得る。

【解決手段】 男性歌唱者の音声である入力音声を変換して得られる女性の音声である元変換音声信号に気息性雑音成分  $S_{ABWN}$  を加算して変換音声信号とし、あるいは、合成した女声の音声である元合成信号に気息性雑音成分を加算して出力音声信号とするので、男声→女声変換あるいは、女声合成を行うに際し、聴感上自然な女性の音声容易に得ることが可能となる。



## 【特許請求の範囲】

【請求項1】 男声の入力音声を変換して出力する音声信号処理装置において、前記入力音声を変換して得られる元変換音声信号に気息性雑音成分信号を加算して前記変換音声信号として出力する気息性雑音加算手段を備えたことを特徴とする音声信号処理装置。

【請求項2】 請求項1記載の音声信号処理装置において、

前記気息性雑音加算手段は、前記入力音声を変換して得られる元変換音声信号のフォルマント周波数を検出するフォルマント周波数検出手段と、

検出した前記フォルマント周波数に対応する周波数帯域を有する前記気息性雑音成分信号を生成する気息性雑音生成手段と、

前記気息性雑音成分信号を前記元変換音声信号に重畳して前記変換音声信号として出力する重畳手段と、を備えたことを特徴とする音声信号処理装置。

【請求項3】 請求項2記載の音声信号処理装置において、

前記気息性雑音生成手段は、ホワイトノイズ信号を生成し出力するホワイトノイズ発生手段と、

前記フォルマント周波数検出手段の検出結果に基づいて、前記ホワイトノイズ信号のうち、前記元変換音声信号の第3フォルマントに対応する所定の周波数帯域成分のみを通過させ元気息性雑音成分信号として出力するバンドパスフィルタ手段と、

前記元変換音声信号に基づいて、前記元気息性雑音成分信号の信号レベルを制御して前記気息性雑音成分信号として出力する信号レベル制御手段と、を備えたことを特徴とする音声信号処理装置。

【請求項4】 女性の音声を作成して出力する音声信号処理装置であって、

前記合成により得られる元合成音声信号に気息性雑音成分信号を加算して前記変換音声信号として出力する気息性雑音加算手段を備えたことを特徴とする音声信号処理装置。

【請求項5】 請求項4記載の音声信号処理装置において、

前記気息性雑音加算手段は、前記元合成音声信号のフォルマント周波数に対応する所定の周波数帯域を有する前記気息性雑音成分信号を生成する気息性雑音生成手段と、

前記気息性雑音成分信号を前記元合成音声信号に重畳して前記変換音声信号として出力する重畳手段と、を備えたことを特徴とする音声信号処理装置。

【請求項6】 請求項5記載の音声信号処理装置において、

前記気息性雑音生成手段は、ホワイトノイズ信号を生成し出力するホワイトノイズ発生手段と、

前記ホワイトノイズ信号のうち、前記元変換音声信号の第3フォルマントに対応する所定の周波数帯域成分のみを通過させ元気息性雑音成分信号として出力するバンドパスフィルタ手段と、

前記元変換音声信号に基づいて、前記元気息性雑音成分信号の信号レベルを制御して前記気息性雑音成分信号として出力する信号レベル制御手段と、を備えたことを特徴とする音声信号処理装置。

【請求項7】 男声の入力音声を変換して出力する音声信号処理方法において、

前記入力音声を変換して得られる元変換音声に気息性雑音成分を加算して前記変換音声とする気息性雑音加算工程を備えたことを特徴とする音声信号処理方法。

【請求項8】 女性の音声を作成する音声信号処理方法であって、

前記合成により得られる元合成音声に気息性雑音成分を加算して前記変換音声とする気息性雑音加算工程を備えたことを特徴とする音声信号処理方法。

## 【発明の詳細な説明】

## 【0001】

【発明の属する技術分野】この発明は、入力音声を変換して出力したり、合成音声を作成する音声信号処理装置及び音声信号処理方法に係り、特に男声→女声変換の機能を有するカラオケ装置に用いるのに好適な音声信号処理装置及び音声信号処理方法に関する。

## 【0002】

【従来の技術】入力された音声の周波数特性などを変えて出力する音声変換装置は種々開発されており、例えば、カラオケ装置の中には、歌手の歌った歌声のピッチを変換して、男性の声を女性の声に変換させるものがある（例えば、特表平8-508581号）。

## 【0003】

【発明が解決しようとする課題】しかしながら、従来の音声変換装置においては、単に歌声のピッチを変換しているだけであるため、男声→女声変換を行っても聴感上自然な女性の音声を得られないという問題点があった。そこで、本発明の目的は、男声→女声の音声変換を行うに際し、聴感上自然な女性の変換音声を容易に得ることが可能な音声信号処理装置及び音声信号処理方法を提供することにある。

## 【0004】

【課題を解決するための手段】上記課題を解決するため、請求項1記載の構成は、男声の入力音声を変換して出力する音声信号処理装置において、前記入力音声を変換して得られる元変換音声信号に気息性雑音成分信号を加算して前記変換音声信号として出力する気息性雑音加算手段を備えたことを特徴としている。

【0005】請求項2記載の構成は、請求項1記載の構成において、前記気息性雑音加算手段は、前記入力音声

を女声変換して得られる元変換音声信号のフォルマント周波数を検出するフォルマント周波数検出手段と、検出した前記フォルマント周波数に対応する周波数帯域を有する前記気息性雑音成分信号を生成する気息性雑音生成手段と、前記気息性雑音成分信号を前記元変換音声信号に重畳して前記変換音声信号として出力する重畳手段と、を備えたことを特徴としている。

【0006】請求項3記載の構成は、請求項2記載の構成において、前記気息性雑音生成手段は、ホワイトノイズ信号を生成し出力するホワイトノイズ発生手段と、前記フォルマント周波数検出手段の検出結果に基づいて、前記ホワイトノイズ信号のうち、前記元変換音声信号の第3フォルマントに対応する所定の周波数帯域成分のみを通過させ元気息性雑音成分信号として出力するバンドパスフィルタ手段と、前記元変換音声信号に基づいて、前記元気息性雑音成分信号の信号レベルを制御して前記気息性雑音成分信号として出力する信号レベル制御手段と、を備えたことを特徴としている。

【0007】請求項4記載の構成は、女性の音声合成して出力する音声信号処理装置であって、前記合成により得られる元合成音声信号に気息性雑音成分信号を加算して前記変換音声信号として出力する気息性雑音加算手段を備えたことを特徴としている。

【0008】請求項5記載の構成は、請求項4記載の構成において、前記気息性雑音加算手段は、前記元合成音声信号のフォルマント周波数に対応する所定の周波数帯域を有する前記気息性雑音成分信号を生成する気息性雑音生成手段と、前記気息性雑音成分信号を前記元合成音声信号に重畳して前記変換音声信号として出力する重畳手段と、を備えたことを特徴としている。

【0009】請求項6記載の構成は、請求項5記載の構成において、前記気息性雑音生成手段は、ホワイトノイズ信号を生成し出力するホワイトノイズ発生手段と、前記ホワイトノイズ信号のうち、前記元変換音声信号の第3フォルマントに対応する所定の周波数帯域成分のみを通過させ元気息性雑音成分信号として出力するバンドパスフィルタ手段と、前記元変換音声信号に基づいて、前記元気息性雑音成分信号の信号レベルを制御して前記気息性雑音成分信号として出力する信号レベル制御手段と、を備えたことを特徴としている。

【0010】請求項7記載の構成は、男声の入力音声を変換する音声信号処理方法において、前記入力音声を変換して得られる元変換音声に気息性雑音成分を加算して前記変換音声とする気息性雑音加算工程を備えたことを特徴としている。

【0011】請求項8記載の構成は、女性の音声合成する音声信号処理方法であって、前記合成により得られる元合成音声に気息性雑音成分を加算して前記変換音声とする気息性雑音加算工程を備えたことを特徴としている。

## 【0012】

【発明の実施の形態】次に図面を参照して本発明の好適な実施形態について説明する。

### 〔1〕 実施形態の原理構成

始めに、実施形態の原理について図15の原理説明図を参照して説明する。〔1.1〕 実施形態の原理構成  
音声信号処理装置100は、マイク101から入力された男声／女声変換して元変換音声信号として出力する男声／女声変換部102と、元変換音声信号のフォルマント周波数の検出を行うフォルマント周波数検出部103と、検出された第3フォルマントの周波数に基づいて、後述するバンドパスフィルタの通過帯域制御を行うバンドパスフィルタ特性制御部104と、ホワイトノイズを発生しホワイトノイズ信号として出力するホワイトノイズ発生部105と、バンドパスフィルタ特性制御部104の制御下で、第3フォルマントに対応する所定の周波数帯域の周波数を有するホワイトノイズ信号のみを元気息性ノイズ信号として通過させるバンドパスフィルタ部106と、元変換音声信号の周波数－アンプ軸上におけるアンプの大きさを検出するアンプ・エンベロープ検出部107と、検出したアンプの大きさに基づいて元気息性ノイズ信号の信号レベルを制御する信号レベル制御信号を出力する気息性雑音レベル制御部108と、信号レベル制御信号に基づいて元気息性ノイズ信号の信号レベルを変化させ、気息性ノイズ信号として出力するアンプ部109と、元変換音声信号に気息性ノイズ信号を加算して変換音声信号として出力するミキサ部110と、変換音声信号に基づいて電気／音響変換を行い音響信号として出力するスピーカ部111と、を備えて構成されている。

### 【0013】〔1.2〕 実施形態の原理動作

次に原理動作を説明する。マイク101から入力された男声は、男声／女声変換部102により、男声→女声変換がなされて元変換音声信号としてフォルマント周波数検出部103及びアンプ・エンベロープ検出部107に出力される。フォルマント周波数検出部103は、元変換音声信号のフォルマント周波数（特に第3フォルマント）の検出を行う。バンドパスフィルタ特性制御部104は、フォルマント周波数検出部103により検出された第3フォルマントの周波数に基づいて、バンドパスフィルタの通過帯域制御を行う。

【0014】これにより、バンドパスフィルタ部106は、バンドパスフィルタ特性制御部104の制御下で、ホワイトノイズ発生部105により出力されたホワイトノイズ信号のうち、第3フォルマントに対応する所定の周波数帯域の周波数を有するホワイトノイズ信号のみを元気息性ノイズ信号として通過させ、アンプ部109に出力する。一方、気息性雑音レベル制御部108は、アンプ・エンベロープ検出部107が検出した元変換音声信号の周波数－アンプ軸上におけるアンプの大きさに基

づいて元気息性ノイズ信号の信号レベルを制御する信号レベル制御信号をアンプ部109に出力する。

【0015】アンプ部109は、信号レベル制御信号に基づいて元気息性ノイズ信号の信号レベルを変化させ、気息性ノイズ信号としてミキサ部110に出力し、ミキサ部110は、元変換音声信号に気息性ノイズ信号を加算して変換音声信号としてスピーカ部111に出力する。そしてスピーカ部111は、変換音声信号に基づいて電気／音響変換を行い音響信号として出力することとなる。

【0016】[1. 3] まとめ

これらの処理において、得られる変換音声信号信号によれば、再生される音響信号（音声）は、元歌唱者の歌声（男声）が、あたかも、女性歌唱者が歌った自然な女性の歌声のようになる。

【0017】[2] 実施形態の詳細構成

図1及び図2に、実施形態の詳細構成図を示す。なお、本実施形態は、本発明による音声変換装置（音声変換方法）をカラオケ装置に適用し、より自然な音声変換を行うことができるカラオケ装置として構成した場合の例である。図1において、マイク1は、元歌唱者（me）の声を収集し、入力音声信号 $S_v$ として入力音声信号切出部3に出力する。これと並行して、分析窓生成部2は、前回のフレームで検出したピッチの周期の固定倍（例えば、3.5倍など）の周期を有する分析窓（例えば、ハミング窓）AWを生成し、入力音声信号切出部3に出力する。なお、初期状態あるいは前回のフレームが無声音（含む無音）の場合には、予め設定した固定周期の分析窓を分析窓AWとして入力音声信号切出部3に出力する。

【0018】これらにより入力音声信号切出部3は、入力された分析窓AWと入力音声信号 $S_v$ とを掛け合わせ、入力音声信号 $S_v$ をフレーム単位で切り出し、フレーム音声信号 $F S_v$ として高速フーリエ変換部4に出力される。より具体的には、入力音声信号 $S_v$ とフレームとの関係は、図3に示すようになっており、各フレーム $F L$ は、前のフレーム $F L$ と一部重なるように設定されている。そして、高速フーリエ変換部4においてフレーム音声信号 $F S_v$ は、解析処理されるとともに、図4に示すように、高速フーリエ変換部4の出力である周波数スペクトルからピーク検出部5によりローカルピークが検出される。

【0019】より具体的には、図4に示すような周波数スペクトルに対して、×印を付けたローカルピークを検出する。このローカルピークは、周波数値とアンプ（振幅）値の組み合わせとして表される。すなわち、図4に示すように、 $(F_0, A_0)$ 、 $(F_1, A_1)$ 、 $(F_2, A_2)$ 、……、 $(F_N, A_N)$ というように各フレームについてローカルピークが検出され、表されることとなる。そして、図3に模式的に示すように、各フレー

ム毎に一組（以下、ローカルピーク組という。）として無声／有声検出部6及びピーク連携部8に出力される。

無声／有声検出部6は、入力されたフレーム毎のローカルピークに基づいて、高周波成分の大きさに応じて無声であることを検出（‘t’、‘k’等）し、無声／有声検出信号 $U/V_{me}$ をピッチ検出部7、イーजीシンクロナイゼーション処理部22及びクロスフェーダ部30に出力する。あるいは、時間軸上で単位時間あたりの零クロス数に応じて無声であることを検出（‘s’等）し、元無声／有声検出信号 $U/V_{me}$ をピッチ検出部7、イー

【0020】さらに無声／有声検出部6は、入力されたフレームが無声であると検出されなかった場合には、入力されたローカルピーク組をそのまま、ピッチ検出部7に出力する。ピッチ検出部7は、入力されたローカルピーク組に基づいて、当該ローカルピーク組が対応するフレームのピッチ $P_{me}$ を検出する。より具体的なフレームのピッチ $P_{me}$ の検出方法としては、例えば、Maher, R.C. and J.W. Beauchamp: "Fundamental Frequency Estimation of Musical Signal using a two-way Mismatch Procedure" (Journal of Acoustical Society of America 95 (4): 2254-2263) に開示されているような方法で行う。

【0021】次に、ピーク検出部5から出力されたローカルピーク組は、ピーク連携部8において、前後のフレームについて連携が判断され、連携すると認められるローカルピークについては、一連のデータ列となるようにローカルピークをつなげる連携処理がなされる。ここで、この連携処理について、図5を参照して説明する。今、図5（A）に示すようなローカルピークが前回のフレームにおいて検出され、図5（B）に示すようなローカルピークが今回のフレームにおいて検出されたとする。

【0022】この場合、ピーク連携部8は、前回のフレームで検出された各ローカルピーク（ $F_0, A_0$ ）、（ $F_1, A_1$ ）、（ $F_2, A_2$ ）、……、（ $F_N, A_N$ ）に対応するローカルピークが今回のフレームでも検出されたか否かを調べる。対応するローカルピークがあるか否かの判断は、前回のフレームで検出されたローカルピークの周波数を中心にした所定範囲内に今回のフレームのローカルピークが検出されるか否かによって行われる。より具体的には、図5の例では、ローカルピーク（ $F_0, A_0$ ）、（ $F_1, A_1$ ）、（ $F_2, A_2$ ）……については、対応するローカルピークが検出されているが、ローカルピーク（ $F_K, A_K$ ）については（図5（A）参照）、対応するローカルピーク（図5（B）参照）は検出されていない。

【0023】ピーク連携部8は、対応するローカルピークを検出した場合は、それらを時系列順に繋げて一組のデータ列として出力する。なお、対応するローカルピー

クが検出されない場合は、当該フレームについての対応ローカルピークは無しということを示すデータに置き換える。ここで、図6は、複数のフレームにわたるローカルピークの周波数 $F_0$ 及び周波数 $F_1$ の変化の一例を示している。このような変化は、アンプ（振幅） $A_0$ 、 $A_1$ 、 $A_2$ 、……についても同様に認められる。この場合、ピーク連携部8から出力されるデータ列は、フレームの間隔おきに出力される離散的な値である。

【0024】なお、ピーク連携部8から出力されるピーク値を、以後において、確定成分という。これは、元の信号（すなわち、音声信号 $S_v$ ）のうち正弦波の要素として確定的に置き換えられる成分という意味である。また、置き換えられた各正弦波（厳密には、正弦波のパラメータである周波数及びアンプ（振幅））の各々については、正弦波成分と呼ぶことにする。

【0025】次に、補間合成部9は、ピーク連携部8から出力される確定成分について補間処理を行い、補間後の確定成分に基づいていわゆるオシレータ方式で波形合成を行う。この場合の補間の間隔は、後述する出力部34が出力する最終出力信号のサンプリングレート（例えば、44.1kHz）に対応した間隔で行われる。前述した図6に示す実線は、正弦波成分の周波数 $F_0$ 、 $F_1$ について補間処理が行われた場合のイメージを示している。

【0026】〔2.1〕 補間合成部の構成  
ここで、補間合成部9の構成を図7に示す。補間合成部9は、複数の部分波形発生部9aを備えて構成されており、各部分波形発生部9aは、指定された正弦波成分の周波数（ $F_0$ 、 $F_1$ 、…）およびアンプ（振幅）に応じた正弦波を発生する。ただし、本第1実施形態における正弦波成分（ $F_0$ 、 $A_0$ ）、（ $F_1$ 、 $A_1$ ）、（ $F_2$ 、 $A_2$ ）、……は、各々補間の間隔に従って時事刻々変化していくものであるから、各部分波形発生部9aから出力される波形は、その変化に従った波形になる。すなわち、ピーク連携部8からは正弦波成分（ $F_0$ 、 $A_0$ ）、（ $F_1$ 、 $A_1$ ）、（ $F_2$ 、 $A_2$ ）、……が順次出力され、各正弦波成分の各々について補間処理が行われるから、各部分波形発生部9aは、所定の周波数領域内で周波数と振幅が変動する波形を出力する。そして、各部分波形発生部9aから出力された波形は、加算部9bにおいて加算合成される。したがって、補間合成部9の出力信号は、入力音声信号 $S_v$ から確定成分を抽出した正弦波成分合成信号 $S_{SS}$ になる。

【0027】〔2.2〕 残差成分検出部の動作  
次に、残差成分検出部10は、補間合成部9から出力された正弦波成分合成信号 $S_{SS}$ と入力音声信号 $S_v$ との偏差である残差成分信号 $S_{RD}$ （時間波形）を生成する。この残差成分信号 $S_{RD}$ は、音声に含まれる無声成分を多く含む。一方、前述の正弦波成分合成信号 $S_{SS}$ は有声成分

る歌唱者の声に似せるには、有声音についてだけ処理を行えば、無声音については処理を施す必要はあまりない。そこで、本実施形態においては、有声音成分に対応する確定成分について音声変換処理を行うようにしている。より具体的には、残差成分信号 $S_{RD}$ については、高速フーリエ変換部11で、周波数波形に変換し、得られた残差成分信号（周波数波形）を $R_{me}(f)$ として残差成分保持部12に保持しておく。

【0028】〔2.3〕 平均アンプ演算部の動作  
一方、図7（A）に示すように、ピーク検出部5からピーク連携部8を介して出力された正弦波成分（ $F_0$ 、 $A_0$ ）、（ $F_1$ 、 $A_1$ ）、（ $F_2$ 、 $A_2$ ）、……、（ $F_N$ 、 $A_N$ ）の $N$ 個の正弦波成分（以下、これらをまとめて $F_n$ 、 $A_n$ と表記する。 $n=0\sim(N-1)$ 。）は、正弦波成分保持部13に保持されるとともに、アンプ $A_n$ は平均アンプ演算部14に入力され、各フレーム毎に次式により平均アンプ $A_{me}$ が算出される。

$$A_{me} = \Sigma(A_n) / N$$

【0029】〔2.4〕 アンプ正規化部の動作  
次にアンプ正規化部15において、次式により各アンプ $A_n$ を平均アンプ $A_{me}$ で正規化し、正規化アンプ $A'_n$ を求める。

$$A'_n = A_n / A_{me}$$

〔2.5〕 スペクトラル・シェイプ演算部の動作  
そして、スペクトラル・シェイプ演算部16において、図8（B）に示すように、周波数 $F_n$ 及び正規化アンプ $A'_n$ により得られる正弦波成分（ $F_n$ 、 $A'_n$ ）をブレークポイントとするエンベロープ（包絡線）をスペクトラル・シェイプ $S_{me}(f)$ として生成する。この場合において、二つのブレークポイント間の周波数におけるアンプの値は、当該二つのブレークポイントを、例えば、直線補間することにより算出する。なお、補間の方法は直線補間に限られるものではない。

【0030】〔2.6〕 ピッチ正規化部の動作  
続いてピッチ正規化部17においては、各周波数 $F_n$ をピッチ検出部7において検出したピッチ $P_{me}$ で正規化し、正規化周波数 $F'_n$ を求める。

$$F'_n = F_n / P_{me}$$

これらの結果、元フレーム情報保持部18は、入力音声信号 $S_v$ に含まれる正弦波成分に対応する元属性データである平均アンプ $A_{me}$ 、ピッチ $P_{me}$ 、スペクトラル・シェイプ $S_{me}(f)$ 、正規化周波数 $F'_n$ を保持することとなる。なお、この場合において、正規化周波数 $F'_n$ は、倍音列の周波数の相対値を表しており、もし、フレームの倍音構造を完全倍音構造であるとして取り扱うならば、保持する必要はない。

【0031】この場合において、男声／女声変換を行おうとしている場合には、この段階において、男声→女声変換を行う場合には、ピッチをオクターブ上げ、女声→男声変換を行う場合にはピッチをオクターブ下げる男声

／女声ピッチ制御処理を行うようにするのが好ましい。つづいて、元フレーム情報保持部18に保持している元属性データのうち、平均アンプ $A_{me}$ およびピッチ $P_{me}$ については、さらに静的変化／ビブラートの変化分離部19により、フィルタリング処理などを行って、静的変化成分とビブラート変化的成分とに分離して保持する。なお、さらにビブラート変化的成分からより高周波変化成分であるジッタ変化的成分を分離するように構成することも可能である。

【0032】より具体的には、平均アンプ $A_{me}$ を平均アンプ静的成分 $A_{me-sta}$ 及び平均アンプビブラートの成分 $A_{me-vib}$ とに分離して保持する。また、ピッチ $P_{me}$ をピッチ静的成分 $P_{me-sta}$ 及びピッチビブラートの成分 $P_{me-vib}$ とに分離して保持する。これらの結果、対応するフレームの元フレーム情報データ $I N F_{me}$ は、図8(C)に示すように、入力音声信号 $S_v$ の正弦波成分に対応する元属性データである平均アンプ静的成分 $A_{me-sta}$ 、平均アンプビブラートの成分 $A_{me-vib}$ 、ピッチ静的成分 $P_{me-sta}$ 、ピッチビブラートの成分 $P_{me-vib}$ 、スペクトラル・シェイプ $S_{me}(f)$ 、正規化周波数 $F'_n$ 及び残差成分 $R_{me}(f)$ の形で保持されることとなる。

【0033】一方、ものまねの対象(target)となる歌唱者に対応するターゲット属性データから構成されるターゲットフレーム情報データ $I N F_{tar}$ は、予め分析されてターゲットフレーム情報保持部20を構成するハードディスクなどに予め保持されている。この場合において、ターゲットフレーム情報データ $I N F_{tar}$ のうち、正弦波成分に対応するターゲット属性データとしては、平均アンプ静的成分 $A_{tar-sta}$ 、平均アンプビブラートの成分 $A_{tar-vib}$ 、ピッチ静的成分 $P_{tar-sta}$ 、ピッチビブラートの成分 $P_{tar-vib}$ 、スペクトラル・シェイプ $S_{tar}(f)$ がある。また、ターゲットフレーム情報データ $I N F_{tar}$ のうち、残差成分に対応するターゲット属性データとしては、残差成分 $R_{tar}(f)$ がある。

【0034】〔2. 7〕キーコントロール／テンポチェンジ部の動作

次にキーコントロール／テンポチェンジ部21は、シーケンサ21からの同期信号 $S_{SYNC}$ に基づいて、ターゲットフレーム情報保持部20から同期信号 $S_{SYNC}$ に対応するフレームのターゲットフレーム情報 $I N F_{tar}$ の読出処理及び読み出したターゲットフレーム情報データ $I N F_{tar}$ を構成するターゲット属性データの補正処理を行うとともに、読み出したターゲットフレーム情報 $I N F_{tar}$ および当該フレームが無声であるか有聲であるかを表すターゲット無声／有聲検出信号 $U/V_{tar}$ を出力する。より具体的には、キーコントロール／テンポチェンジ部21の図示しないキーコントロールユニットは、カラオケ装置のキーを基準より上げ下げした場合、ターゲット属性データであるピッチ静的成分 $P_{tar-sta}$ 及びピッチビブラートの成分 $P_{tar-vib}$ についても、同じだけ

上げ下げする補正処理を行う。例えば、50 [cent] だけキーを上げた場合には、ピッチ静的成分 $P_{tar-sta}$ 及びピッチビブラートの成分 $P_{tar-vib}$ についても50 [cent] だけ上げなければならない。

【0035】また、キーコントロール／テンポチェンジ部21の図示しないテンポチェンジユニットは、カラオケ装置のテンポを上げ下げした場合には、変更後のテンポに相当するタイミングで、ターゲットフレーム情報データ $I N F_{tar}$ の読み出し処理を行う必要がある。この場合において、必要なフレームに対応するタイミングに相当するターゲットフレーム情報データ $I N F_{tar}$ が存在しない場合には、当該必要なフレームのタイミングの前後のタイミングに存在する二つのフレームのターゲットフレーム情報データ $I N F_{tar}$ を読み出し、これら二つのターゲットフレーム情報データ $I N F_{tar}$ により補間処理を行い、当該必要なタイミングにおけるフレームのターゲットフレーム情報データ $I N F_{tar}$ 、ひいては、ターゲット属性データを生成する。この場合において、ビブラートの成分(平均アンプビブラートの成分 $A_{tar-vib}$ 及びピッチビブラートの成分 $P_{tar-vib}$ )に関しては、そのままでは、ビブラートの周期自体が変化してしまい、不適当であるので、周期が変動しないような補間処理を行う必要がある。又は、ターゲット属性データとして、ビブラートの軌跡そのものを表すデータではなく、ビブラート周期及びビブラート深さのパラメータを保持し、実際の軌跡を演算により求めるようにすれば、この不具合を回避することができる。

【0036】〔2. 8〕イージーシンクロナイゼーション処理部の動作

次にイージーシンクロナイゼーション処理部22は、ものまねをしようとする歌唱者のフレーム(以下、元フレームという。)に元フレーム情報データ $I N F_{me}$ が存在するにもかかわらず、対応するものまねの対象となる歌唱者のフレーム(以下、ターゲットフレームという。)にターゲットフレーム情報データ $I N F_{tar}$ が存在しない場合には、当該ターゲットフレームの前後方向に存在するフレームのターゲットフレーム情報データ $I N F_{tar}$ を当該ターゲットフレームのターゲットフレーム情報データ $I N F_{tar}$ とするイージーシンクロナイゼーション処理を行う。

【0037】そして、イージーシンクロナイゼーション処理部22は、後述する置換済ターゲットフレーム情報データ $I N F_{tar-sync}$ に含まれるターゲット属性データのうち正弦波成分に関するターゲット属性データ(平均アンプ静的成分 $A_{tar-sync-sta}$ 、平均アンプビブラートの成分 $A_{tar-sync-vib}$ 、ピッチ静的成分 $P_{tar-sync-sta}$ 、ピッチビブラートの成分 $P_{tar-sync-vib}$ 及びスペクトラル・シェイプ $S_{tar-sync}(f)$ )を正弦波成分属性データ選択部23に出力する。また、イージーシンクロナイゼーション処理部22は、後述する置換済ターゲット

フレーム情報データ I N F tar-syncに含まれるターゲット属性データのうち残差成分に関するターゲット属性データ（残差成分 R tar-sync(f)）を残差成分選択部 2 5 に出力する。

【0038】このイーजीシンクロナイゼーション部 2 2 における処理においても、ビブラートの成分（平均アンプビブラートの成分 A tar-vib及びピッチビブラートの成分 P tar-vib）に関しては、そのままでは、ビブラートの周期自体が変化してしまい、不適当であるので、周期が変動しないような補間処理を行う必要がある。又は、ターゲット属性データとして、ビブラートの軌跡そのものを表すデータではなく、ビブラート周期及びビブラート深さのパラメータを保持し、実際の軌跡を演算により求めるようにすれば、この不具合を回避することができる。

【0039】[2. 8. 1] イーजीシンクロナイゼーション処理の詳細

ここで、図 9 及び図 10 を参照してイーजीシンクロナイゼーション処理について詳細に説明する。図 9 は、イーजीシンクロナイゼーション処理のタイミングチャートであり、図 10 はイーजीシンクロナイゼーション処理フローチャートである。まず、イーजीシンクロナイゼーション部 2 2 は、シンクロナイゼーション処理の処理方法を表すシンクロナイゼーションモード = “0” とする（ステップ S 1 1）。このシンクロナイゼーションモード = “0” は、元フレームに対応するターゲットフレームにターゲットフレーム情報データ I N F tarが存在する通常処理の場合に相当する。そしてあるタイミング t における元無声／有声検出信号 U / V me(t) が無声（U）から有声（V）に変化したか否かを判別する（ステップ S 1 2）。

【0040】例えば、図 9 に示すように、タイミング t = t 1 においては、元無声／有声検出信号 U / V me(t) が無声（U）から有声（V）に変化している。ステップ S 1 2 の判別において、元無声／有声検出信号 U / V me(t) が無声（U）から有声（V）に変化している場合には（ステップ S 1 2 ; Y e s）、タイミング t の前回のタイミング t -1 における元無声／有声検出信号 U / V me(t-1) が無声（U）かつターゲット無声／有声検出信号 U / V tar(t-1) が無声（U）であるか否かを判別する（ステップ S 1 8）。例えば、図 9 に示すように、タイミング t = t 0 (= t 1-1) においては、元無声／有声検出信号 U / V me(t-1) が無声（U）かつターゲット無声／有声検出信号 U / V tar(t-1) が無声（U）となっている。ステップ S 1 8 の判別において、元無声／有声検出信号 U / V me(t-1) が無声（U）かつターゲット無声／有声検出信号 U / V tar(t-1) が無声（U）となっている場合には（ステップ S 1 8 ; Y e s）、当該ターゲットフレームには、ターゲットフレーム情報データ I N F tarが存在しないので、シンクロナイゼーションモード =

“1” とし、置換用のターゲットフレーム情報データ I N F holdを当該ターゲットフレームの後方向（Backward）に存在するフレームのターゲットフレーム情報とする。

【0041】例えば、図 9 に示すように、タイミング t = t 1 ~ t 2 のターゲットフレームには、ターゲットフレーム情報データ I N F tarが存在しないので、シンクロナイゼーションモード = “1” とし、置換用ターゲットフレーム情報データ I N F holdを当該ターゲットフレームの後方向に存在するフレーム（すなわち、タイミング t = t 2 ~ t 3 に存在するフレーム）のターゲットフレーム情報データ backward とする。そして、処理をステップ S 1 5 に移行し、シンクロナイゼーションモード = “0” であるか否かを判別する（ステップ S 1 5）。ステップ S 1 5 の判別において、シンクロナイゼーションモード = “0” である場合には、タイミング t における元フレームに対応するターゲットフレームにターゲットフレーム情報データ I N F tar(t) が存在する場合、すなわち、通常処理であるので、置換済ターゲットフレーム情報データ I N F tar-syncをターゲットフレーム情報データ I N F tar(t) とする。

I N F tar-sync = I N F tar(t)

【0042】例えば、図 9 に示すようにタイミング t = t 2 ~ t 3 のターゲットフレームには、ターゲットフレーム情報データ I N F tarが存在するので、

I N F tar-sync = I N F tar(t)

とする。この場合において、以降の処理に用いられる置換済ターゲットフレーム情報データ I N F tar-syncに含まれるターゲット属性データ（平均アンプ静的成分 A tar-sync-sta、平均アンプビブラートの成分 A tar-sync-vib、ピッチ静的成分 P tar-sync-sta、ピッチビブラートの成分 P tar-sync-vib、スペクトラル・シェイプ S tar-sync(f) 及び残差成分 R tar-sync(f)）は実質的には、以下の内容となる（ステップ S 1 6）。

A tar-sync-sta = A tar-sta

A tar-sync-vib = A tar-vib

P tar-sync-sta = P tar-sta

P tar-sync-vib = P tar-vib

S tar-sync(f) = S tar(f)

R tar-sync(f) = R tar(f)

【0043】ステップ S 1 5 の判別において、シンクロナイゼーションモード = “1” またはシンクロナイゼーションモード = “1” である場合には、タイミング t における元フレームに対応するターゲットフレームにターゲットフレーム情報データ I N F tar(t) が存在しない場合であるので、置換済ターゲットフレーム情報データ I N F tar-syncを置換用ターゲットフレーム情報データ I N F holdとする。I N F tar-sync = I N F hold例えば、図 9 に示すように、タイミング t = t 1 ~ t 2 のターゲットフレームには、ターゲットフレーム情報データ I N F

tarが存在せず、シンクロナイゼーションモード＝  
 “1”となるが、タイミング $t = t_2 \sim t_3$ のターゲット  
 フレームには、ターゲットフレーム情報データINF tar  
 rが存在するので、置換済ターゲットフレーム情報デー  
 タINF tar-syncをタイミング $t = t_2 \sim t_3$ のターゲッ  
 トフレームのターゲットフレーム情報データである置換  
 用ターゲットフレーム情報データINF holdとする処理  
 P1を行い、以降の処理に用いられる置換済ターゲット  
 フレーム情報データINF tar-syncに含まれるターゲッ  
 ト属性データは、平均アンプ静的成分A tar-sync-sta、  
 平均アンプビブラト的成分A tar-sync-vib、ピッチ静  
 的成分P tar-sync-sta、ピッチビブラト的成分P tar-  
 sync-vib、スペクトラル・シェイプS tar-sync(f)及び  
 残差成分R tar-sync(f)となる（ステップS16）。

【0044】また、図9に示すように、タイミング $t =$   
 $t_3 \sim t_4$ のターゲットフレームには、ターゲットフレ  
 ーム情報データINF tarが存在せず、シンクロナイゼー  
 ションモード＝“2”となるが、タイミング $t = t_2 \sim$   
 $t_3$ のターゲットフレームには、ターゲットフレーム情  
 報データINF tarが存在するので、置換済ターゲット  
 フレーム情報データINF tar-syncをタイミング $t = t_2 \sim$   
 $t_3$ のターゲットフレームのターゲットフレーム情報  
 データである置換用ターゲットフレーム情報データINF  
 holdとする処理P2を行い、以降の処理に用いられる  
 置換済ターゲットフレーム情報データINF tar-syncに  
 含まれるターゲット属性データは、平均アンプ静的成分  
 A tar-sync-sta、平均アンプビブラト的成分A tar-sy-  
 nc-vib、ピッチ静的成分P tar-sync-sta、ピッチビブ  
 ラト的成分P tar-sync-vib、スペクトラル・シェイプS  
 tar-sync(f)及び残差成分R tar-sync(f)となる（ステッ  
 プS16）。ステップS12の判別において、元無声/  
 有声検出信号U/V<sub>me</sub>(t)が無声(U)から有声(V)  
 に変化していない場合には（ステップS12；No）、  
 ターゲット無声/有声検出信号U/V tar(t)が有声  
 (V)から無声(U)に変化しているか否かを判別する  
 （ステップS13）。

【0045】ステップS13の判別において、ターゲッ  
 ト無声/有声検出信号U/V tar(t)が有声(V)から無  
 声(U)に変化している場合には（ステップS13；Ye  
 s）、タイミングtの前のタイミングt-1における  
 元無声/有声検出信号U/V<sub>me</sub>(t-1)が有声(V)かつ  
 ターゲット無声/有声検出信号U/V tar(t-1)が有声  
 (V)であるか否かを判別する（ステップS19）。例  
 えば、図9に示すように、タイミングt3においてター  
 ゲット無声/有声検出信号U/V tar(t)が有声(V)か  
 ら無声(U)に変化し、タイミングt-1=t2～t3にお  
 いては、元無声/有声検出信号U/V<sub>me</sub>(t-1)が有声  
 (V)かつターゲット無声/有声検出信号U/V tar(t-  
 1)が有声(U)となっている。

【0046】ステップS18の判別において、元無声/

有声検出信号U/V<sub>me</sub>(t-1)が有声(V)かつターゲッ  
 ト無声/有声検出信号U/V tar(t-1)が有声(V)とな  
 っている場合には（ステップS19；Yes）、当該ター  
 ゲットフレームには、ターゲットフレーム情報データ  
 INF tarが存在しないので、シンクロナイゼーション  
 モード＝“2”とし、置換用のターゲットフレーム情報  
 データINF holdを当該ターゲットフレームの前方向  
 (forward)に存在するフレームのターゲットフレーム  
 情報とする。例えば、図9に示すように、タイミング $t =$   
 $t_3 \sim t_4$ のターゲットフレームには、ターゲットフレ  
 ーム情報データINF tarが存在しないので、シンクロ  
 ナイゼーションモード＝“2”とし、置換用ターゲッ  
 トフレーム情報データINF holdを当該ターゲットフレ  
 ームの前方向に存在するフレーム（すなわち、タイミン  
 グ $t = t_2 \sim t_3$ に存在するフレーム）のターゲットフレ  
 ーム情報データforwardとする。

【0047】そして、処理をステップS15に移行し、  
 シンクロナイゼーションモード＝“0”であるか否かを  
 判別して（ステップS15）、以下、同様の処理を行  
 う。ステップS13の判別において、ターゲット無声/  
 有声検出信号U/V tar(t)が有声(V)から無声(U)  
 に変化していない場合には（ステップS13；No）、  
 タイミングtにおける元無声/有声検出信号U/V<sub>me</sub>  
 (t)が有声(V)から無声(U)に変化し、あるいは、  
 ターゲット無声/有声検出信号U/V tar(t)が無声  
 (U)から有声(V)に変化しているか否かを判別する  
 （ステップS14）。

【0048】ステップS14の判別において、タイミン  
 グtにおける元無声/有声検出信号U/V<sub>me</sub>(t)が有声  
 (V)から無声(U)に変化し、あるいは、ターゲット  
 無声/有声検出信号U/V tar(t)が無声(U)から有声  
 (V)に変化している場合には（ステップS14；Ye  
 s）、シンクロナイゼーションモード＝“0”とし、置  
 換用ターゲットフレーム情報データINF holdを初期化  
 (clear)し、処理をステップS15に移行して、以  
 下、同様の処理を行う。ステップS14の判別におい  
 て、タイミングtにおける元無声/有声検出信号U/V  
 me(t)が有声(V)から無声(U)に変化せず、あるい  
 は、ターゲット無声/有声検出信号U/V tar(t)が無声  
 (U)から有声(V)に変化していない場合には（ステ  
 ップS14；No）、そのまま処理をステップS15に  
 移行し、以下同様の処理を行う。

【0049】〔2. 9〕 変形スペクトラルシェイプ生  
 成部の動作

続いて、変形スペクトラルシェイプ生成部23は、静的  
 変化/ビブラト的成分分離部19から入力された入力  
 音声信号S<sub>v</sub>の正弦波成分に対応する元属性データであ  
 る平均アンプ静的成分A<sub>me</sub>-sta、平均アンプビブラト  
 的成分A<sub>me</sub>-vib、ピッチ静的成分P<sub>me</sub>-sta、ピッチビブ  
 ラト的成分P<sub>me</sub>-vib、スペクトラル・シェイプS<sub>me</sub>

(f)、正規化周波数 $F'_n$ 、イージーシンクロナイゼーション部22から入力された置換済ターゲットフレーム情報データINF tar-syncに含まれるターゲット属性データのうち正弦波成分に関するターゲット属性データ

(平均アンプ静的成分A tar-sync-sta、平均アンプビブラートの成分A tar-sync-vib、ピッチ静的成分P tar-sync-sta、ピッチビブラートの成分P tar-sync-vib及びスペクトラル・シェイプS tar-sync(f))及びコントローラ29から入力される変形スペクトラル・シェイプ生成情報に基づいて、新しいスペクトラル・シェイプである変形スペクトラル・シェイプS new(f)を生成する。

【0050】変形スペクトラルシェイプの生成は、元歌唱者に対応するスペクトラル・シェイプ(あるいは、ターゲット歌唱者に対応するターゲットスペクトラル・シェイプ)を周波数軸方向に定数 $\alpha$  ( $0 < \alpha \leq 2$ )でシフトすることにより行う。ここで、より具体的に、変形スペクトラルシェイプS new(f)の生成について説明する。

【0051】[2. 9. 1] 変形スペクトラルシェイプS new(f)の具体的生成方法

図11にターゲット歌唱者である女性のスペクトラル・シェイプを示す。図11に示すように、ターゲット歌唱者の正弦波成分に含まれる周波数成分は、 $f_{f0} \sim f_{fn}$ で表されている。図12に元歌唱者である男性のスペクトラル・シェイプを示す。図12に示すように、元歌唱者の正弦波成分に含まれる周波数成分は、 $f_{m0} \sim f_{mn}$ で表されている。また、各周波数成分 $f_{m0} \sim f_{mn}$ に対応するアンプは、 $A_{fm0} \sim A_{fmn}$ で表されている。

【0052】この場合において、元歌唱者のアンプ $A(f_m) = A_{fm0}, A_{fm1}, \dots, A_{fmn}$ は元のままで、周波数成分 $f_{m0} \sim f_{mn}$ のみを $\alpha$ 倍 ( $1 \leq \alpha \leq 2$ )して、すなわち、 $\alpha$ の値に相当するだけスペクトラル・シェイプを周波数軸に沿って高域側にシフトすることにより変形スペクトラル・シェイプS new(f)を生成する。すなわち、変形スペクトラル・シェイプに対応する周波数成分を $f_{h0} \sim f_{hn}$ と表すとすると、

$$f_{h0} = \alpha \cdot f_{m0}$$

$$f_{h1} = \alpha \cdot f_{m1}$$

$$f_{h2} = \alpha \cdot f_{m2}$$

.....

$$f_{hn} = \alpha \cdot f_{mn}$$

とし、図13及び以下に示す変形正弦波成分群(=周波数成分及びアンプで表される正弦波成分の一群)により特定される変形スペクトラルシェイプS new(f)を得る。

( $f_{h0}, A_{fm0}$ )

( $f_{h1}, A_{fm1}$ )

( $f_{h2}, A_{fm2}$ )

.....

( $f_{h0}, A_{fm0}$ )

【0053】ところで、一般的にアンプ成分が大きい場合には、高域まで伸びた抜けの良い音となり、アンプ成

分が小さい場合には、逆にこもった音になる。そこで、新規スペクトラル・シェイプS new(f)に関しては、このような状態をシミュレートすべく、図15に示すように、スペクトラル・シェイプの高域成分、すなわち、高域成分部分のスペクトラル・シェイプの傾きを新規アンプ成分A newの大きさに応じて補償するスペクトラルチルト補償(spectral tilt correction)を行って、コントロールすることにより、よりリアルな音声を再生することができる。続いて、生成された変形スペクトラル・シェイプS new(f)について、必要に応じてコントローラ29から入力される変形スペクトラル・シェイプ加工情報に基づいて、変形スペクトラル・シェイプ加工部24によりさらなる波形の加工を行う。例えば、変形スペクトラル・シェイプS new(f)を全体的に間延びさせる等の波形加工を行う。そして、変形スペクトラル・シェイプ加工部24は得られた変形スペクトラル・シェイプS new(f)に基づいて第3フォルマントを検出する。

【0054】[2. 10] 第3フォルマントFT3の検出

次に、第3フォルマントFT3の検出方法を、隣り合う二つの正弦波成分に対応する正規化アンプ $A'_{fK}, A'_{fK-1}$ の差 $\Delta A'$  ( $f_K - f_{K-1}$ )の変化に基づいて行う場合について説明する。変形スペクトラル・シェイプの第3フォルマントFT3近傍及び第2フォルマントFT2終端部分の状態を図16に示す。第3フォルマントFT3の周波数の範囲は、通常1.5 [kHz]以上4 [kHz]以下であるため、

$$f_K \approx 1.5 \text{ [kHz]}$$

を満たすKを定め、その値をKSとする。また、

$$f_K \approx 4 \text{ [kHz]}$$

を満たすKを定め、その値をKEとする。

【0055】次に、 $f_K$  ( $K=KS, \dots, KE$ )のKの値を増加させる。そして、 $\Delta A'$  ( $f_K - f_{K-1}$ )の値を観察し、

$$\Delta A' (f_K - f_{K-1}) < 0$$

の状態から

$$\Delta A' (f_K - f_{K-1}) \geq 0$$

に変化したときのKの値をpとする。これを $K=KE$ まで繰り返し行って、第3フォルマントFT3の平均周波数に最も近い周波数 $f_p$ を第3フォルマントFT3とすることで検出することができる。なお、第3フォルマントFT3の検出は、上記方法に限られるものではなく、例えば、線形予測法を利用して求めることも可能である。そして、この検出した第3フォルマントFT3は、バンドパスフィルタ特性制御部42に出力される。また、この第3フォルマントFT3の出力と並行して、第3フォルマントFT3の周波数におけるアンプA FT3が検出され、レベル制御部43に出力される。

【0056】[2. 11] 残差成分選択部の動作  
一方、残差成分選択部25は、イージーシンクロナイゼ

ーション部22から入力された置換済ターゲットフレーム情報データINF tar-syncに含まれるターゲット属性データのうち残差成分に関するターゲット属性データ

(残差成分 $R_{tar-sync}(f)$ )、残差成分保持部12に保持されている残差成分信号(周波数波形) $R_{me}(f)$ 及びコントローラ29から入力される残差成分属性データ選択情報に基づいて新しい残差成分属性データである新規残差成分 $R_{new}(f)$ を生成する。すなわち、新規残差成分 $R_{new}(f)$ については、次式により生成する。

$R_{new}(f) = R^*(f)$  (ただし、\*は、 $me$ 又は $tar-sync$ )

この場合においては、 $me$ 又は $tar-sync$ のいずれを選択するかは、新規スペクトラル・シェイプ $S_{new}(f)$ と同一のものを選択するのがより好ましい。

【0057】さらに、新規残差成分 $R_{new}(f)$ に関しても、新規スペクトラル・シェイプと同様な状態をシミュレートすべく、図10に示したように、残差成分の高域成分、すなわち、高域成分部分の残差成分の傾きを新規アンプ成分 $A_{new}$ の大きさに応じて補償するスペクトラルチルト補償(spectral tilt correction)を行って、コントロールすることにより、よりリアルな音声を再生することができる。

【0058】[2.12] 正弦波成分生成部の動作  
続いて、正弦波成分生成部26は、変形スペクトラル・シェイプ加工部24から出力された波形加工を伴わない、あるいは、波形加工を伴う変形スペクトラル・シェイプ $S_{new}(f)$ に基づいて、当該フレームにおける新たな正弦波成分( $F''_0, A''_0$ )、( $F''_1, A''_1$ )、( $F''_2, A''_2$ )、……、( $F''_{(N-1)}, A''_{(N-1)}$ )のN個の正弦波成分(以下、これらをまとめて $F''_n, A''_n$ と表記する。 $n=0 \sim (N-1)$ 。)を求める。より具体的には、周波数成分 $X$ における変形スペクトラル・シェイプ $S_{new}(f)$ のアンプを $A(X)$ と表すとする

と、各正弦波成分( $F''_0, A''_0$ )、( $F''_1, A''_1$ )、( $F''_2, A''_2$ )、……、( $F''_{(N-1)}, A''_{(N-1)}$ )は以下のように表すことができる。

( $F''_0, A''_0$ ) = ( $f_{f0}, A(f_{f0})$ )

( $F''_1, A''_1$ ) = ( $f_{f1}, A(f_{f1})$ )

( $F''_2, A''_2$ ) = ( $f_{f2}, A(f_{f2})$ )

……

( $F''_{(N-1)}, A''_{(N-1)}$ ) = ( $f_{fn}, A(f_{fn})$ )

【0059】[2.13] 正弦波成分変形部の動作  
さらに、求めた新規周波数 $f''_n$ および新規アンプ $a''_n$ について、必要に応じてコントローラ29から入力される正弦波成分変形情報に基づいて、正弦波成分変形部27によりさらなる変形を行う。例えば、偶数倍音成分の新規アンプ $A''_n (= A''_0, A''_2, A''_4, \dots)$ だけを大きく(例えば、2倍する)等の変形を行う。これによって得られる変換音声にさらにバラエティを持たせることが可能となる。

【0060】[2.14] 逆高速フーリエ変換部の動

作

次に逆高速フーリエ変換部28は、求めた新規周波数 $f''_n$ および新規アンプ $A''_n (= \text{新規正弦波成分})$ 並びに新規残差成分 $R_{new}(f)$ をFFTバッファに格納し、順次逆FFTを行い、さらに得られた時間軸信号を一部重複するようにオーバーラップ処理し、それらを加算する加算処理を行うことにより新しい有声音の時間軸信号である変換音声信号を生成する。このとき、コントローラ29から入力される正弦波成分/残差成分バランス制御信号に基づいて、正弦波成分及び残差成分の混合比率を制御し、よりリアルな有声音を得る。この場合において、一般的には、残差成分の混合比率を大きくするとざらついた声を得られる。

【0061】この場合において、FFTバッファに新規周波数 $f''_n$ および新規アンプ $a''_n (= \text{新規正弦波成分})$ 並びに新規残差成分 $R_{new}(f)$ を格納するに際し、異なるピッチ、かつ、適当なピッチで変換された正弦波成分をさらに加えることにより変換音声信号としてハーモニーを得ることができる。さらにシーケンサ31により伴奏音に適合したハーモニーピッチを与えることにより、伴奏に適合した音楽的ハーモニーを得ることができる。

【0062】[2.15] クロスフェーダの動作  
次にクロスフェーダ30は、元無声/有声検出信号 $U/V_{me}(t)$ に基づいて、入力音声信号 $S_v$ が無声( $U$ )である場合には、入力音声信号 $S_v$ をそのままミキサ30に出力する。また、入力音声信号 $S_v$ が有声( $V$ )である場合には、逆FFT変換部28が出力した変換音声信号をミキサ30に出力する。この場合において、切替スイッチとしてクロスフェーダ30を用いているのは、クロスフェード動作を行わせることによりスイッチ切替時のクリック音の発生を防止するためである。

【0063】[2.16] シーケンサ及び音源部の動作

一方、シーケンサ31は、カラオケの伴奏音を発生するための音源制御情報を例えば、MIDI (Musical Instrument Digital Interface) データなどとして音源部32に出力する。これにより音源部32は、音源制御情報に基づいて伴奏信号を生成し、ミキサ33に出力する。

[2.17] バンドパスフィルタ特性制御部、ホワイトノイズ発生部及びバンドパスフィルタの動作  
バンドパスフィルタ特性制御部42は、変形スペクトラル・シェイプ加工部24から出力された第3フォルマント $FT3$ に基づいて、バンドパスフィルタ(BPF)41の通過帯域を第3フォルマント $FT3$ の近傍の周波数帯域とすべく、特性制御信号 $SB_C$ をバンドパスフィルタ41に出力する。

【0064】一方、ホワイトノイズ発生部40は、ホワイトノイズ信号 $S_{WN}$ を生成し、バンドパスフィルタ41に出力する。バンドパスフィルタ41は、バンドパスフ

フィルタ特性制御部42の制御下で、ホワイトノイズ信号SWNのうち、第3フォルマントFT3に対応する所定の周波数帯域の周波数を有するホワイトノイズ信号SWNのみを元気息性ノイズ信号SBWNとして通過させ、アンプ部44に出力する。

【0065】[2.18] レベル制御部及びアンプ部の動作

一方、レベル制御部43は、変形スペクトラル・シェイプ加工部24から出力されたアンプAFT3に基づいて、元気息性ノイズ信号SBWNの信号レベルを制御するための信号レベル制御信号SLCをアンプ部44に出力する。アンプ部44は、信号レベル制御信号SLCに基づいて元気息性ノイズ信号SBWNの信号レベルを変化させ、気息性ノイズ信号SABWNとしてミキサ33に出力することとなる。

[2.19]、ミキサ及び出力部

ミキサ33は、入力音声信号Svあるいは変換音声信号のいずれか一方、気息性ノイズ信号SABWN及び伴奏信号を混合し、混合信号を出力部34に出力する。出力部34は、図示しない増幅器を有し混合信号を増幅して音響信号として出力することとなる。

【0066】[3] 実施形態の変形例

[3.1] 第1変形例

上記第1実施形態（特に図15）においては、男性の音声を変換する音声信号処理装置について説明したが、女性の音声を合成することが可能な女声合成装置についても適用が可能である。この場合において、第1実施形態の場合には、第3フォルマントの周波数を検出していたが、合成後に第3フォルマントの周波数を検出する構成に限らず、予め第3フォルマントの周波数を設定し、これらの情報を記憶しておくように構成することも可能である。

【0067】[3.2] 第2変形例

上記実施形態の説明においては、元歌唱者のスペクトラル・シェイプに基づいて生成した変形スペクトラル・シェイプ及びターゲット歌唱者のターゲット音声信号の正弦波成分に含まれる周波数成分に基づいて正弦波成分群を算出し、変換音声を得る構成としていたが、ターゲット歌唱者のスペクトラル・シェイプに基づいて生成した変形スペクトラル・シェイプ及び元歌唱者の入力音声信号の正弦波成分に含まれる周波数成分に基づいて正弦波成分群を算出し、変換音声を得る構成とすることも可能である。

【0068】[3.3] 第3変形例

正弦波成分の抽出は、この実施形態で用いた方法に限らない。要は、音声信号に含まれる正弦波成分を抽出できればよい。

【0069】[3.4] 第4変形例

本実施形態においては、ターゲットの正弦波成分及び残差成分を記憶したが、これに換えて、ターゲットの音声

そのものを記憶し、それを読み出してリアルタイム処理によって正弦波成分と残差成分とを抽出してもよい。すなわち、本実施形態でものまねをしようとする歌唱者の音声に対して行った処理と同様の処理をターゲットの歌唱者の音声に対して行ってもよい。

【0070】[4] 実施形態の効果

以上の結果、カラオケの伴奏とともに、元歌唱者である男性歌唱者の歌が変換されて出力されることとなるが、得られる変換音声は、聴感上自然な女性の音声ものとなる。

【0071】

【発明の効果】以上説明したように、この発明によれば、男声→女声変換を行うに際し、聴感上自然な女性の変換音声を容易に得ることが可能となる。

【図面の簡単な説明】

【図1】 本発明の一実施形態の構成を示すブロック図（その1）である。

【図2】 本発明の一実施形態の構成を示すブロック図（その2）である。

【図3】 実施形態におけるフレームの状態を示す図である。

【図4】 実施形態における周波数スペクトルのピーク検出を説明するための説明図である。

【図5】 実施形態におけるフレーム毎のピーク値の連携を示す図である。

【図6】 実施形態における周波数値の変化状態を示す図である。

【図7】 実施形態における処理過程における確定成分の変化状態を示す図である。

【図8】 実施形態における信号処理の説明図である。

【図9】 イージーシンクロナイゼーション処理のタイミングチャートである。

【図10】 イージーシンクロナイゼーション処理フローチャートである。

【図11】 女性のスペクトラル・シェイプを説明する図である。

【図12】 男性のスペクトラル・シェイプを説明する図である。

【図13】 男声→女声変換の処理説明図である。

【図14】 スペクトラル・シェイプのスペクトラルチルト補償について説明する図である。

【図15】 実施形態の原理説明図である。

【図16】 第3フォルマントの検出方法の説明図である。

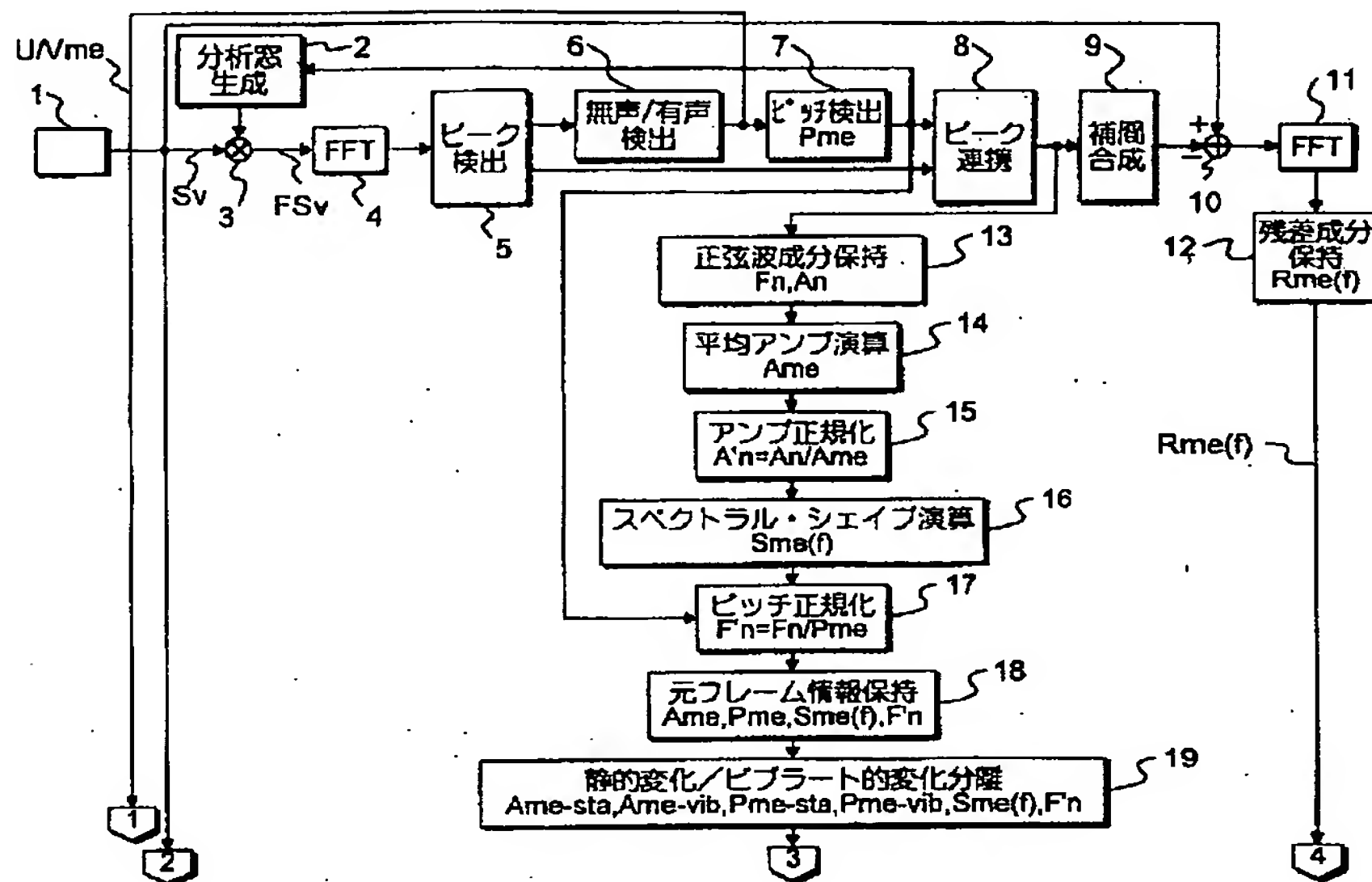
【符号の説明】

1…マイク、2…分析窓生成部、3…入力音声信号切出部、4…高速フーリエ変換部、5…ピーク検出部、6…無声／有声検出部、7…ピッチ抽出部、8…ピーク連携部、9…補間合成部、10…残差成分検出部、11…高速フーリエ変換部、12…残差成分保持部、13…正弦

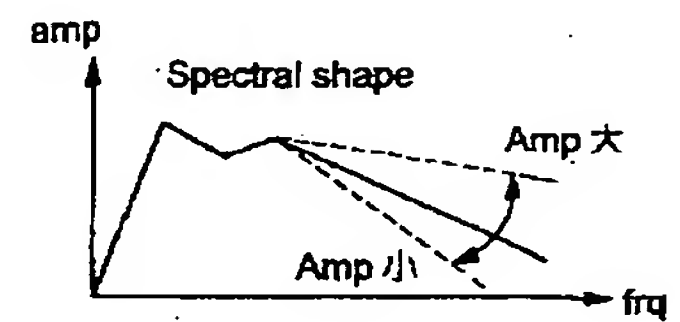
波成分保持部、14…平均アンプ演算部、15…アンプ正規化部、16…スペクトラル・シェイプ演算部、17…ピッチ正規化部、18…元フレーム情報保持部、19…静的変化／ビブラートの变化分離部、20…ターゲットフレーム情報保持部、21…キーコントロール／テンポチェンジ部、22…イージング・シンクロナイズ処理部、23…変形スペクトラル・シェイプ生成部、24…変形スペクトラル・シェイプ加工部、25…残差成分

分選択部、26…正弦波成分生成部、27…正弦波成分変形部、28…逆高速フーリエ変換部、29…コントローラ、30…クロスフェーダ部、31…シーケンサ、32…音源部、33…ミキサ、34…出力部、40…ホワイトノイズ発生部、41…バンドパスフィルタ（BPF）、42…バンドパスフィルタ特性制御部、43…レベル制御部、44…アンプ部、AFT3…第3フォルマントアンプ、FT3…第3フォルマント（周波数）

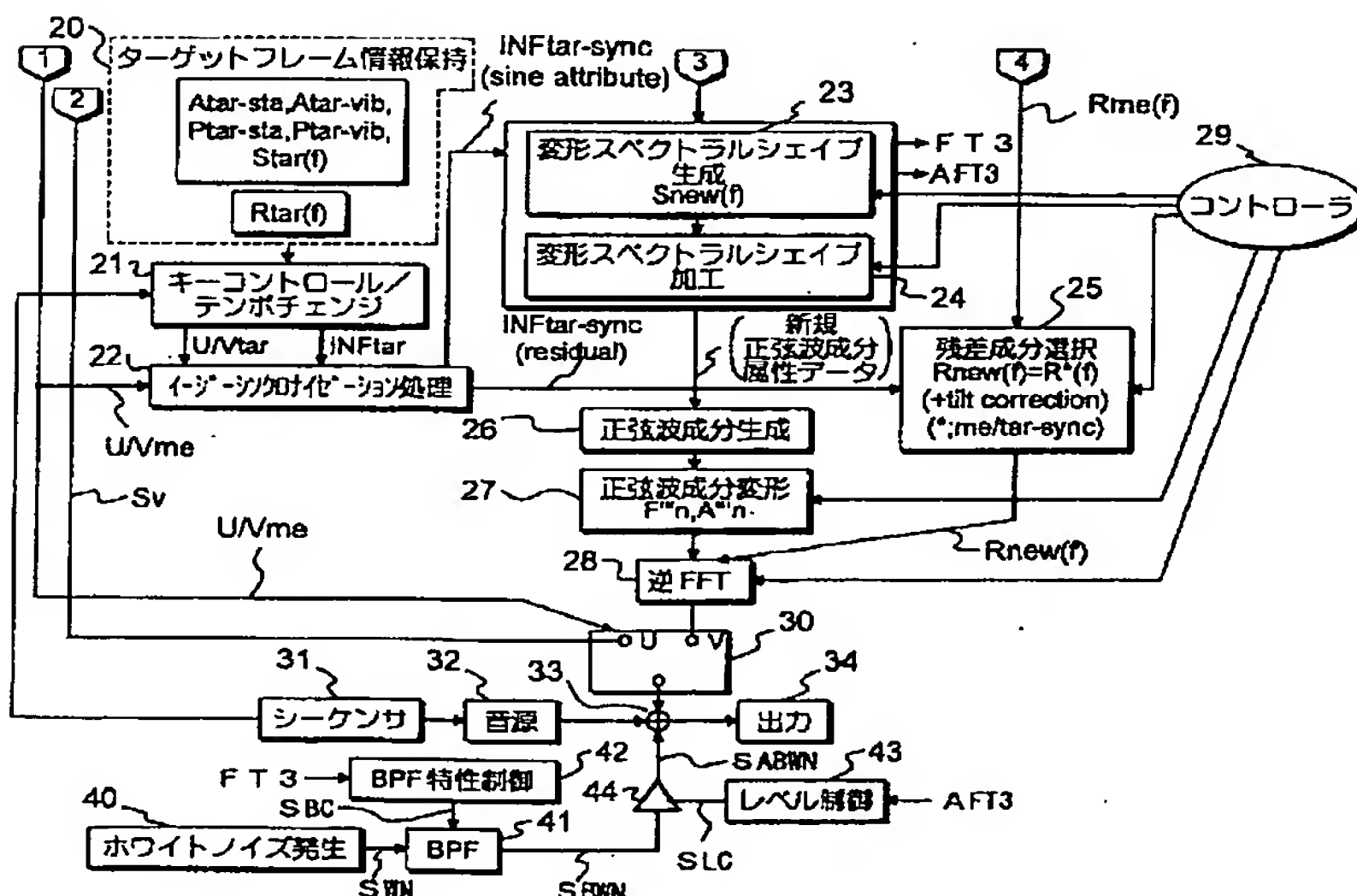
【図1】



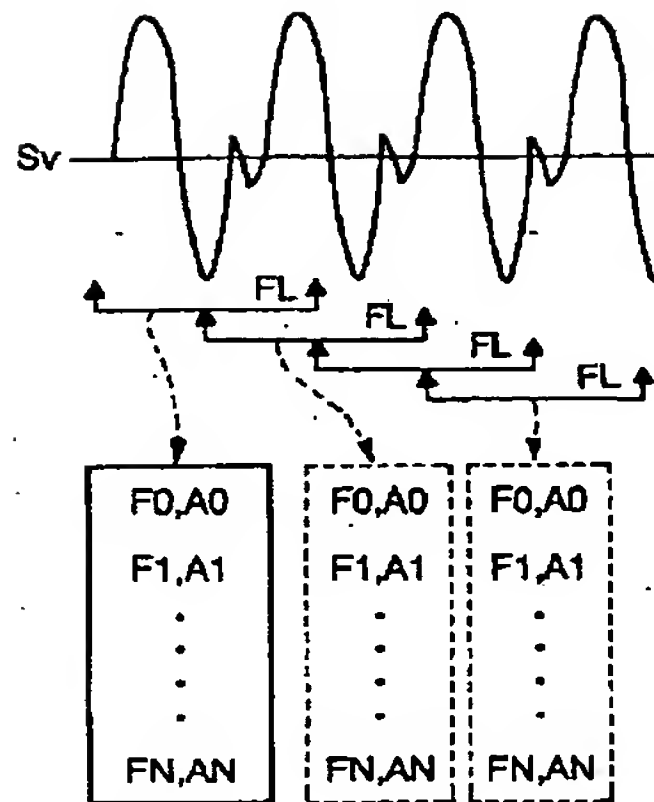
【図14】



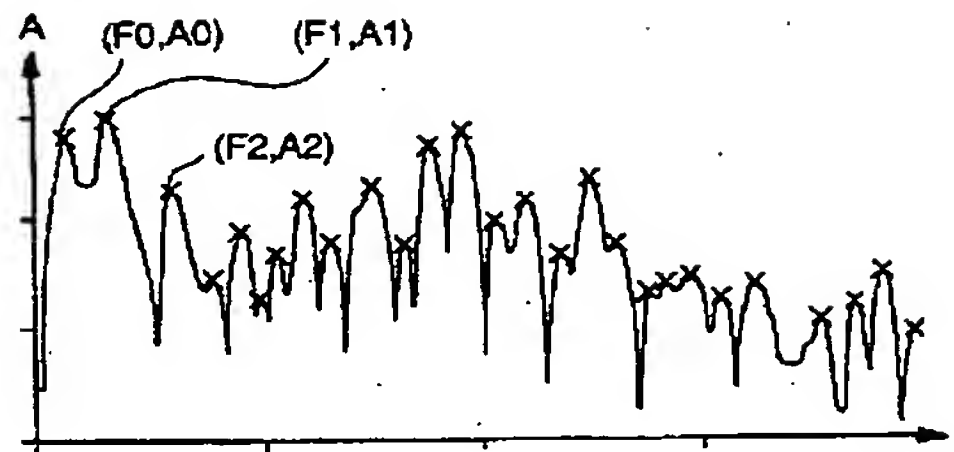
【図2】



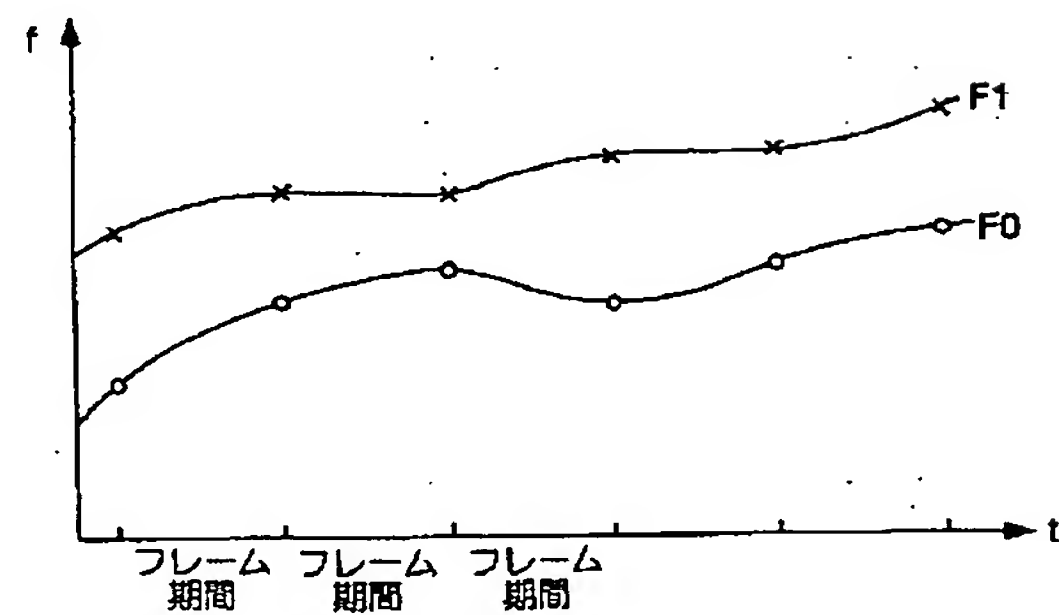
【図3】



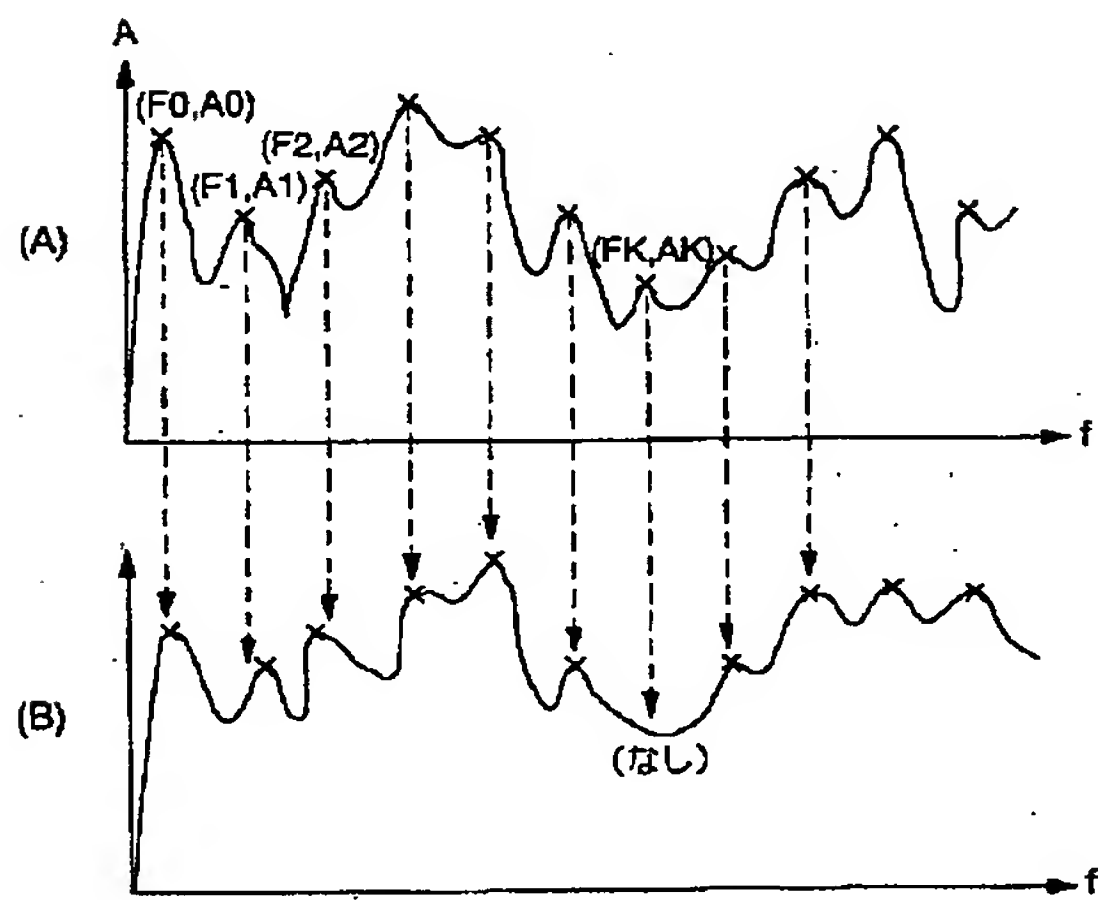
【図4】



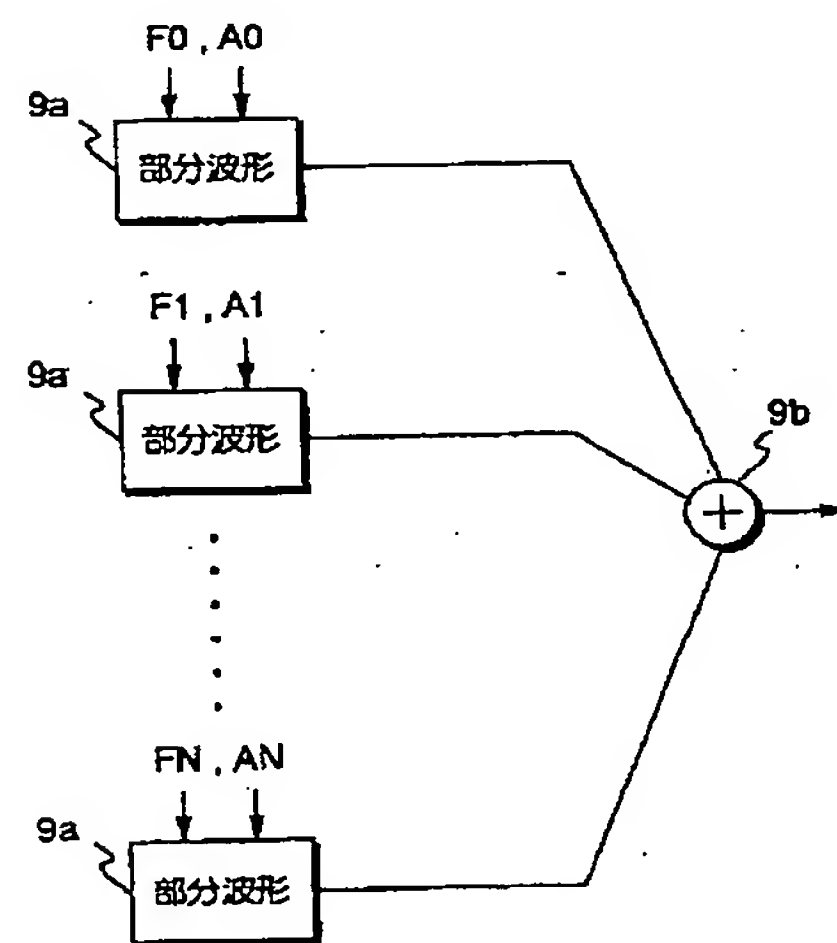
【図6】



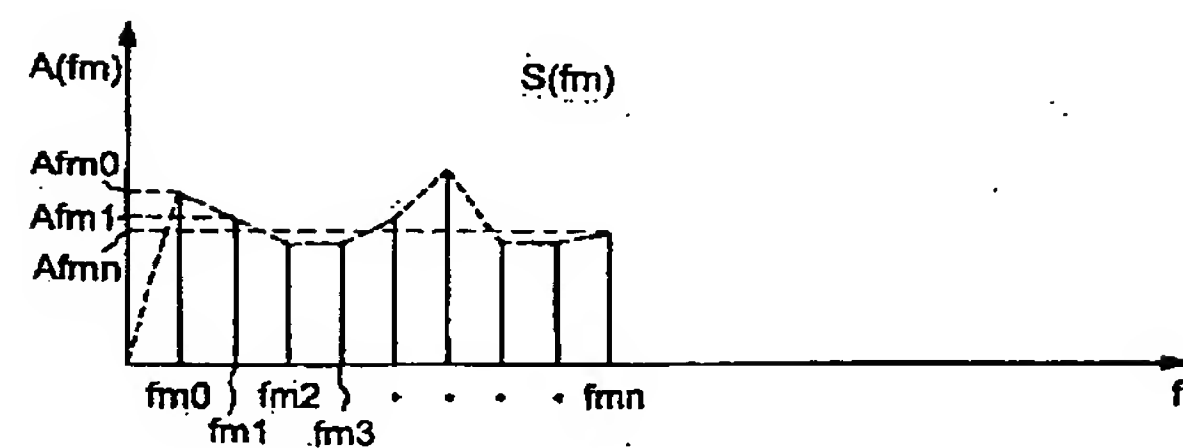
【図5】



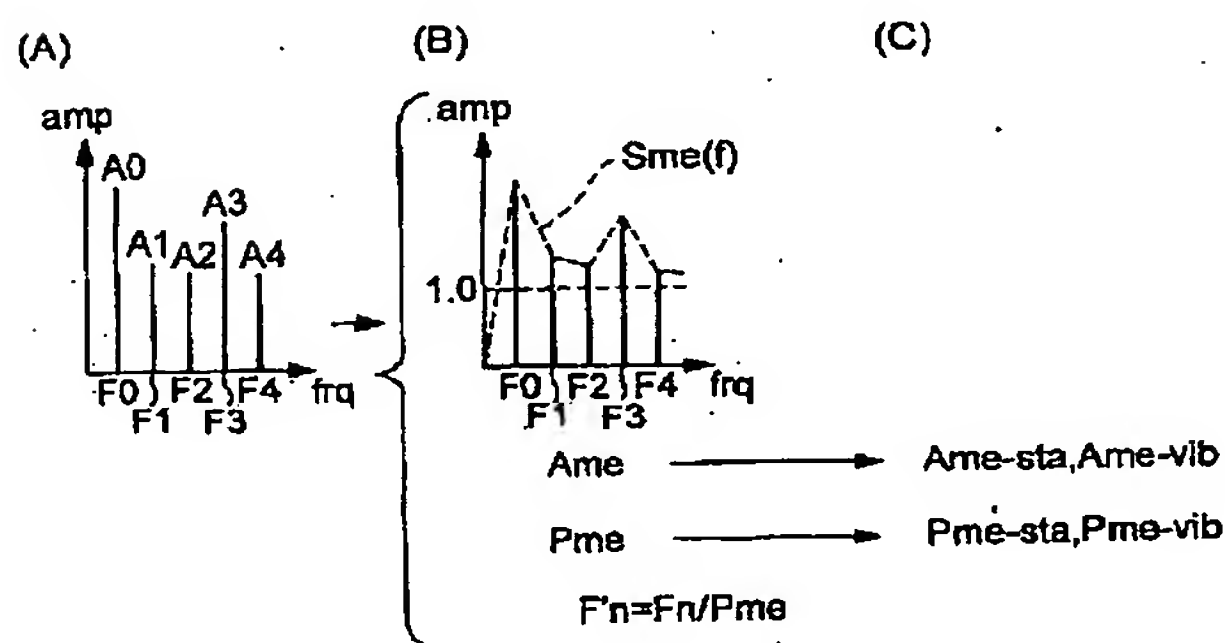
【図8】



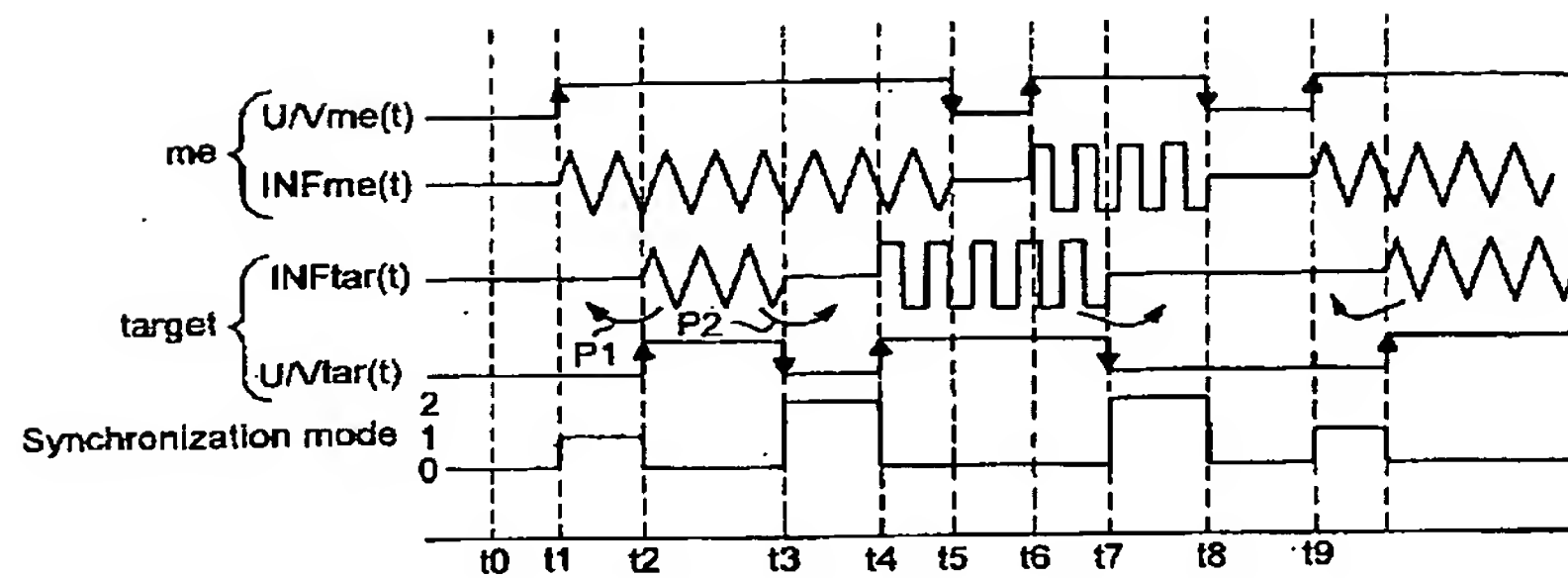
【図12】



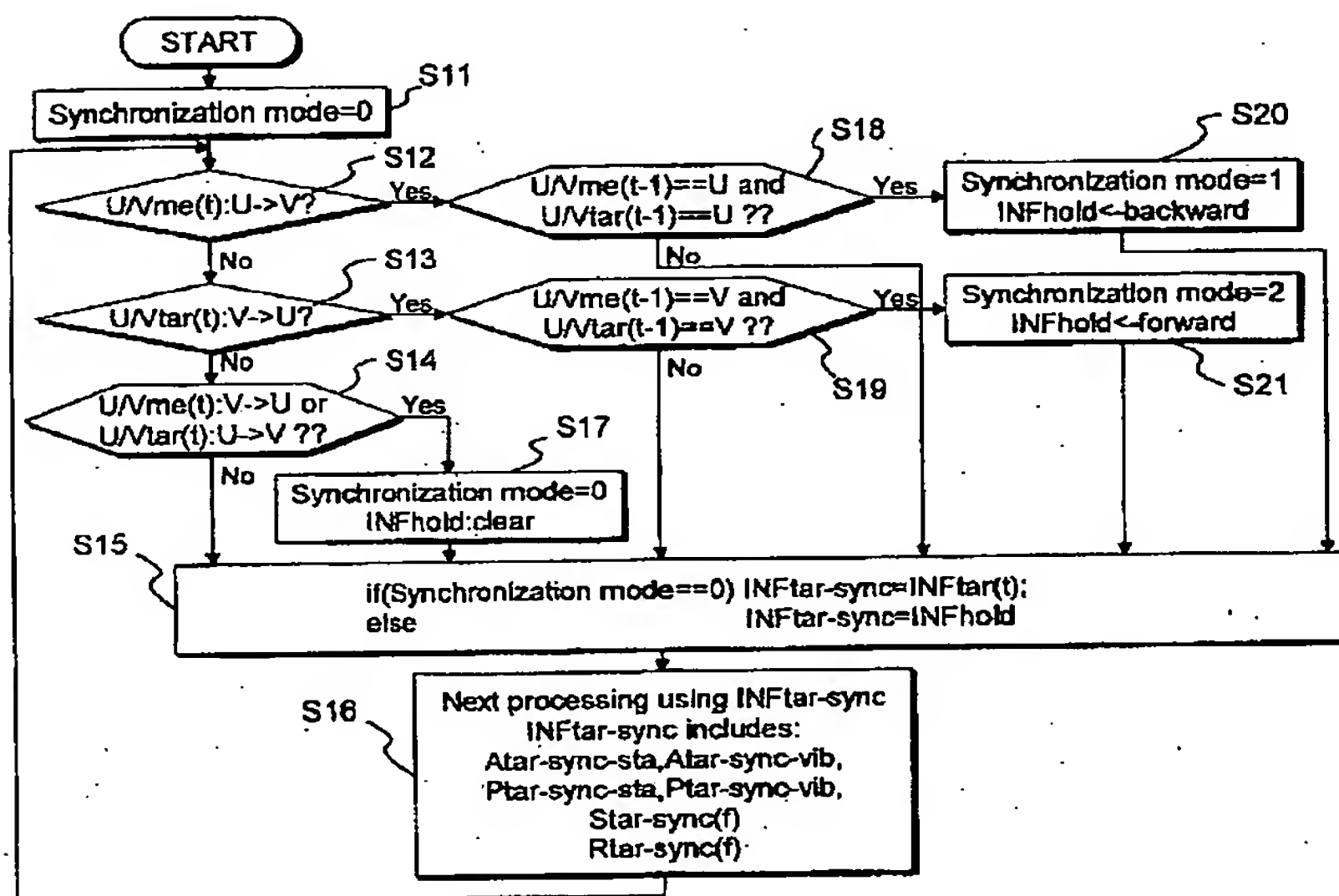
【図7】



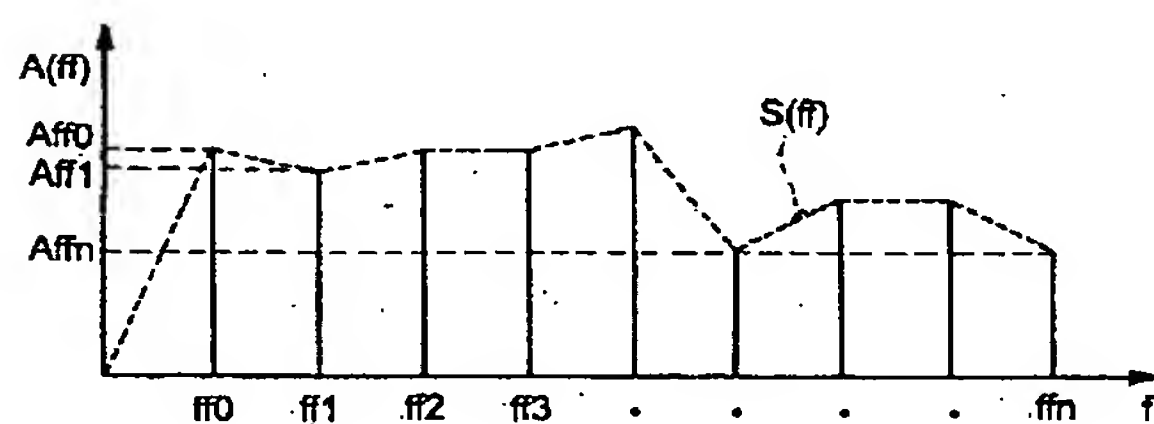
【図9】



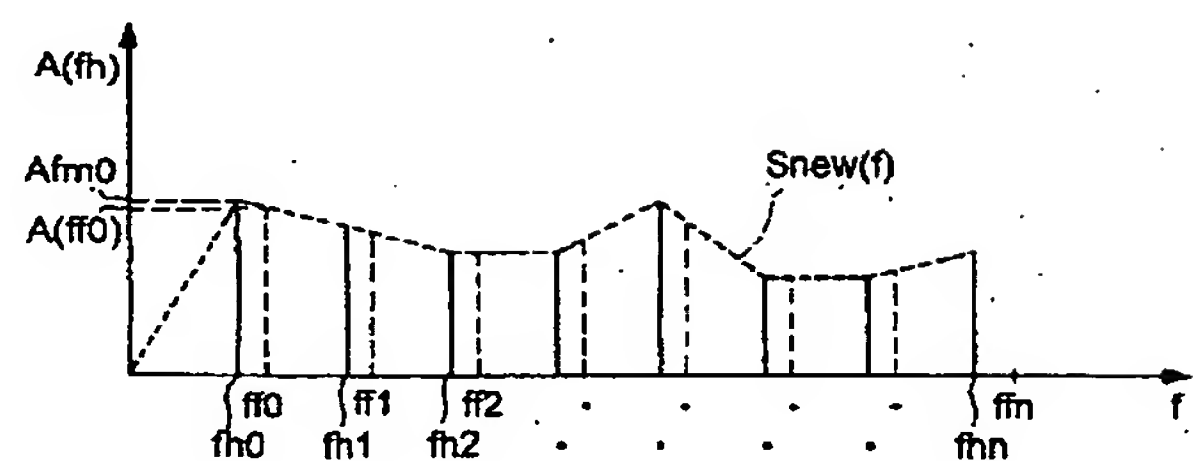
【図10】



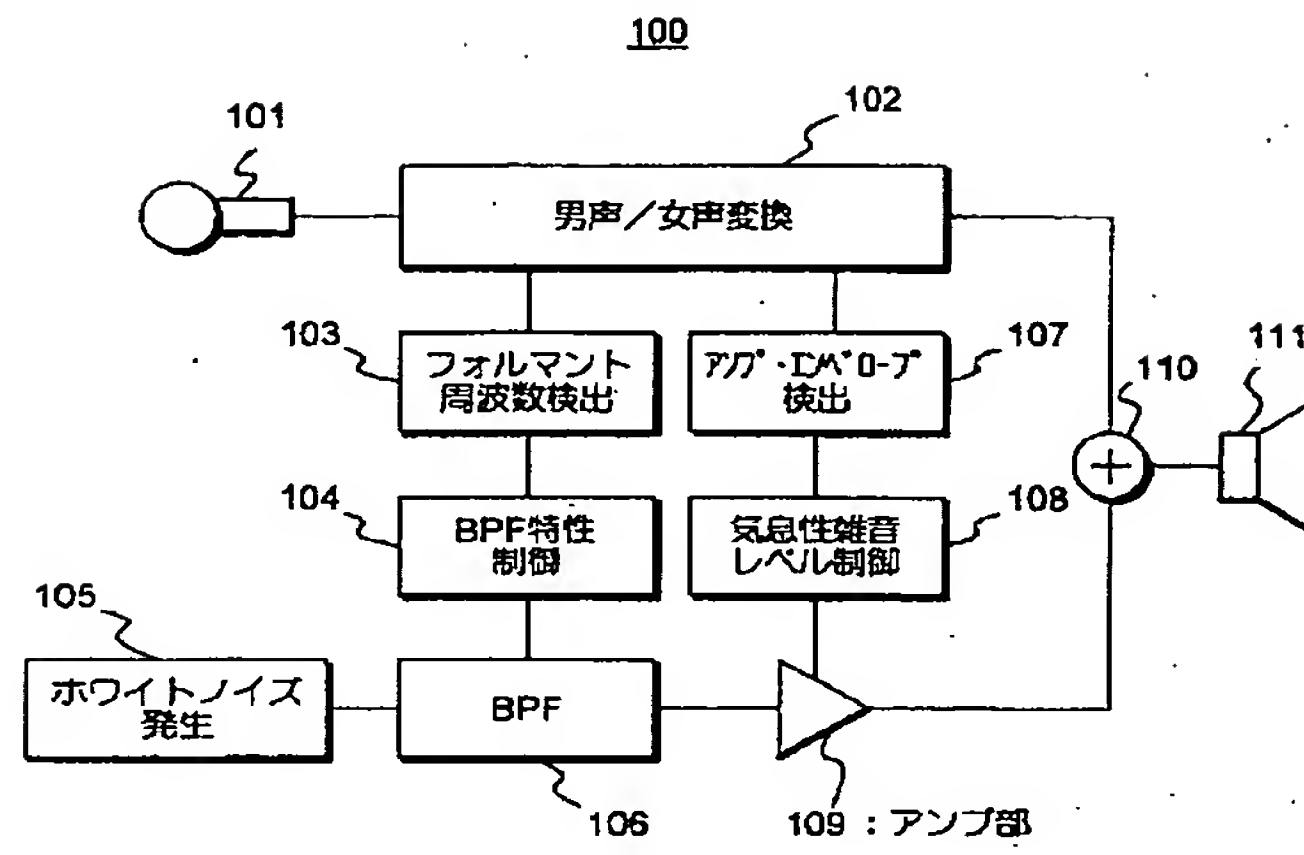
【図11】



【図13】



【図15】



【図16】

